Network Data from ISPs: Uses and Privacy Risks

Nick Feamster
Princeton University

Summary

- ISPs, developers, and researchers use network data for research, development, security.
- Each type of data poses certain privacy risks.
- We need to understand how each type of data is used, and what the risks are, to make appropriate recommendations.

Questions

- What data do ISPs collect?
- How do ISPs use this data, for security and performance?
- What are the privacy risks to individuals?
- When might it be shared? (a. Third parties who provide security services, network management tools. b. Researchers.)
- What are the concerns about privacy rules? (Note: usefulness of data to researchers. Many studies we do would not be possible without data sharing.)

Let's Assume that the Following is True

- End-to-end encryption protects certain aspects of user privacy.
  - Yet, much can be learned about users from encrypted traffic, too.
  - Websites are not the only destinations users are visiting

- DPI is not pervasively deployed
  - Even if it were, retention and analysis is not easy

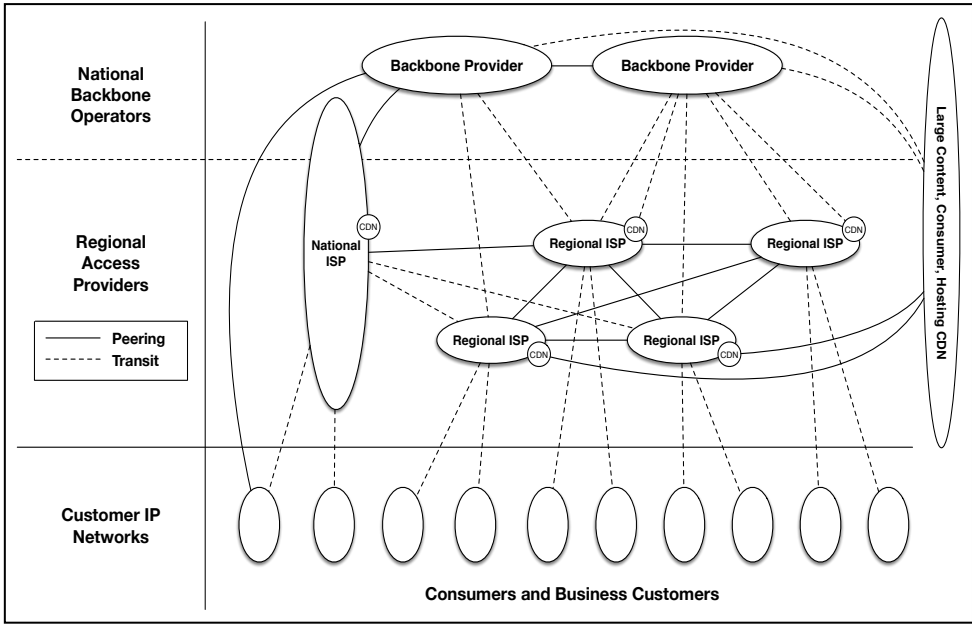
Yet, even if we assume these statements are true,
there are plenty of other types of network data to consider…

Many Other Types of Network Traffic

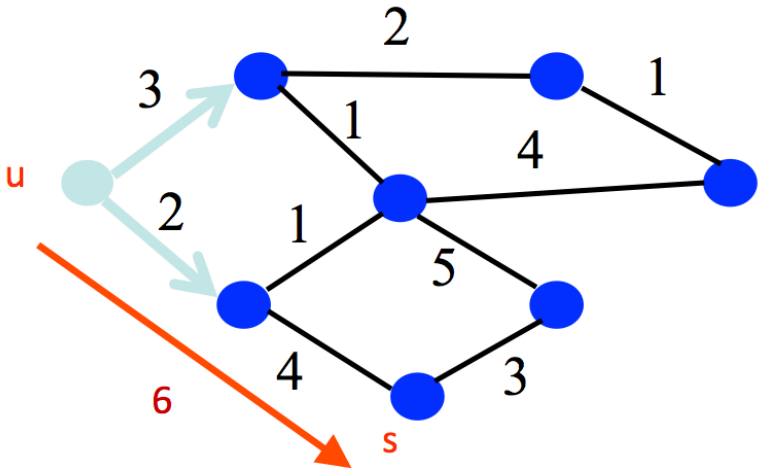
- Routing Data
  - BGP
  - IGP
- Traffic Data
  - NetFlow
  - DNS
  - Simple Network Management Protocol

Routing Data: What is It, How Used?

- BGP Routing
  - How ISPs connect
  - How ISPs choose routes between one another



- IGP Routing
  - Internal network topology



Routing Data: Privacy Risks

- None to invidividuals…
- Routing data reveals the network topology and ISP business decisions, but nothing about user behavior.

Data About Traffic

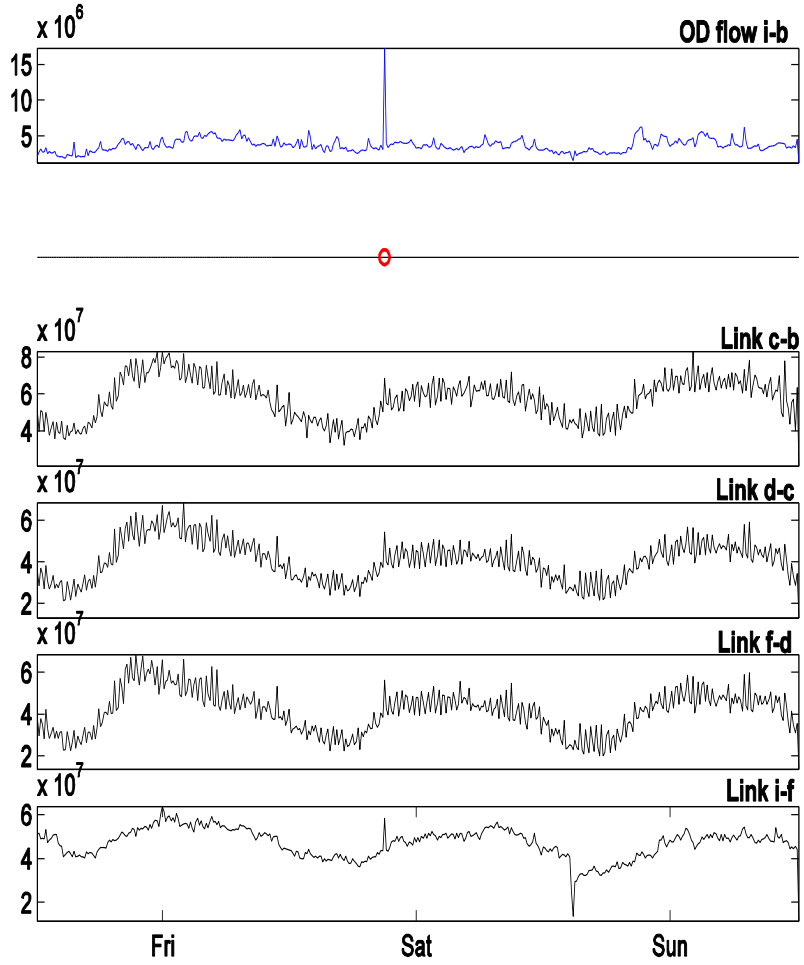
- NetFlow/IPFIX (“Metadata”)
- DNS Lookups
- Simple Network Management Prtocol

NetFlow/IPFIX: What is It, How Used?

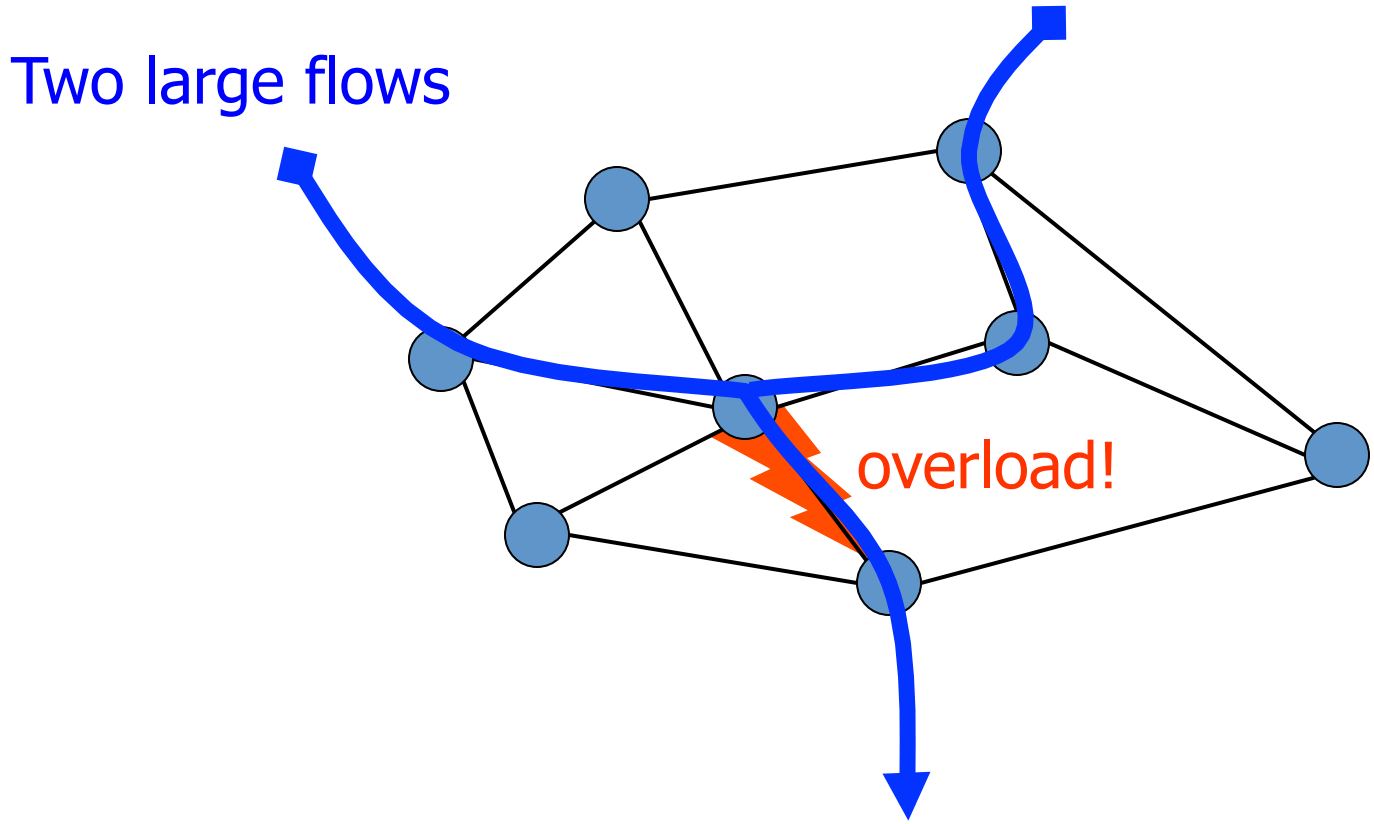
- Security
  - Detection of malware, botnets anomalies
- Performance
  - Traffic engineering
  - Provisioning and planning

Security: Anomaly Detection

- IPFIX data gives information about traffic volumes
- Sudden changes in traffic volume can reveal attacks, failures, or other anomalous activity



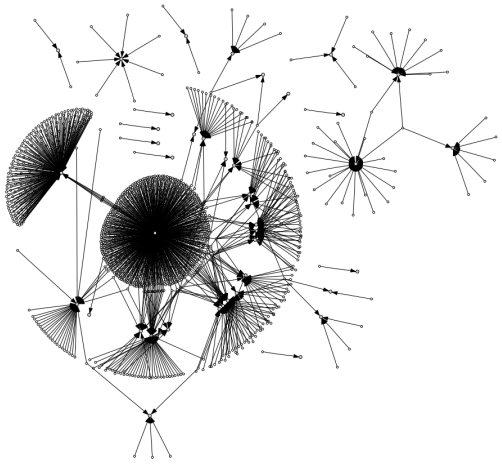
Performance: Traffic Engineering



IPFIX: When Shared, and With Whom?

- With third-party security vendors
  - Detection and mitigation of Denial of Service (DoS)

- With researchers
  - Development of new algorithms to improve performance and security
  - Better understanding of user behavior

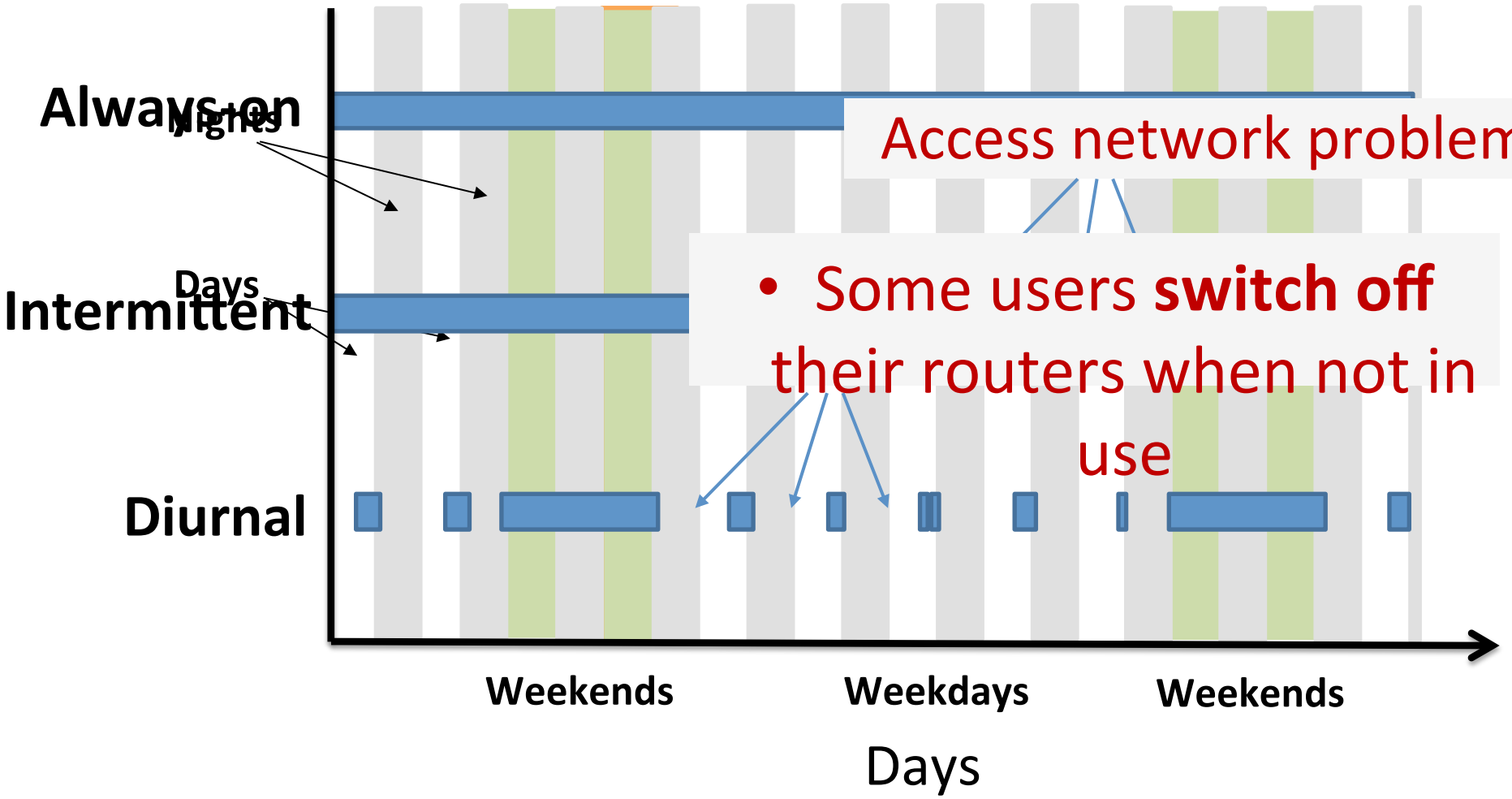


Risks to User Privacy

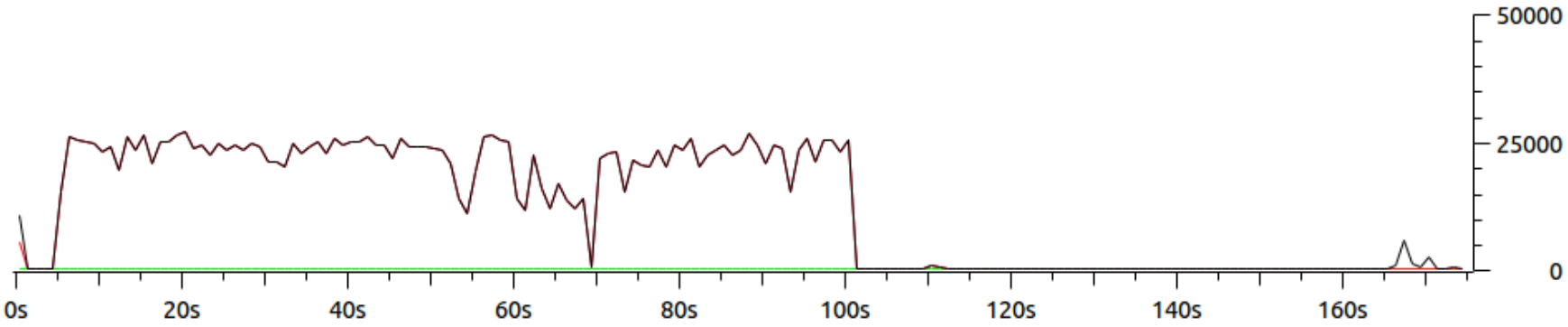
- Traffic volumes can reveal usage patterns
  - Which IP destinations the user visits
  - Levels of network activity
  - What application is the user using?

- Human Behavior
  - At home?
  - Awake or asleep?
  - Which devices?

Example of Activity Patterns

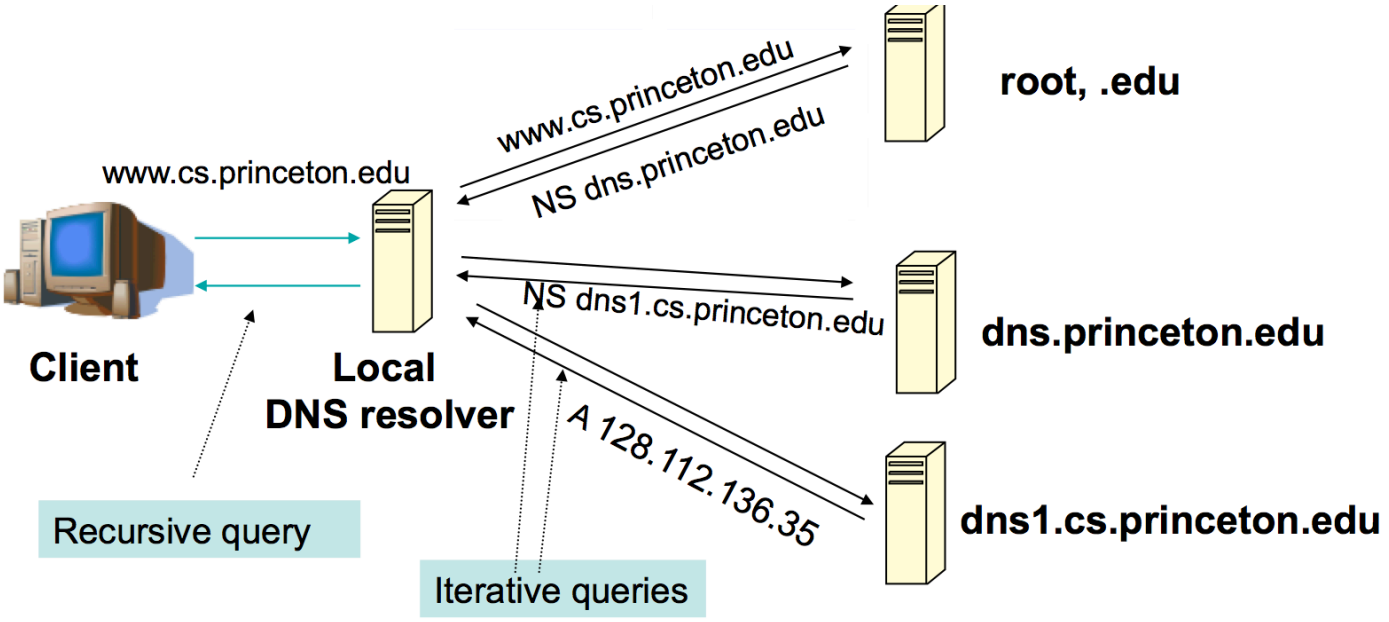


IoT Example: Smartthings Outlet



- Activity patterns reflected in changes in network traffic volume.
- The types of activities are also evident.
- *Works even on encrypted traffic*

DNS: What is it, How Used?



- Maps domain names to IP addresses
- Domain name system (DNS) queries and responses reveal the sites that a user is visiting.
  - What is visible about user depends on vantage point
- For the most part, completely unencrypted

DNS: How Is it Used, When is it Shared?

- Performance
  - Load balancing traffic between different server replicas
- Security
  - Monitoring lookups and domain name registration can often reveal malicious behavior
- Lookups and responses shared with researchers, third parties

DNS: Privacy Risks

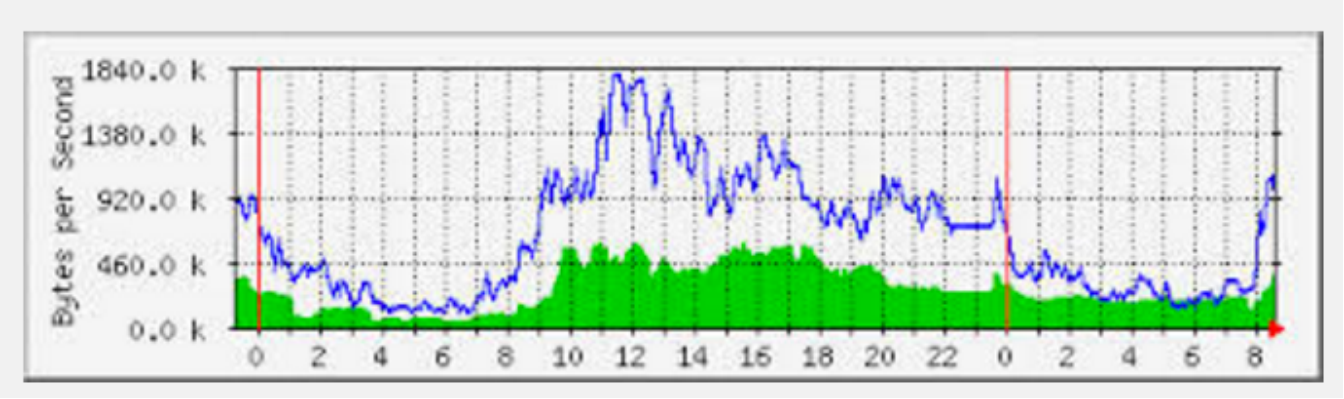
- Can determine:
  - Which sites a user is visiting
  - Which devices users have in their home
  - Activities and interactions with individual devices
- Example from user home

TOP DOMAINS VISITED

Domain	Usage (MB)	Usage (%)
akamai.net	3932.3	31%
google.com	2965.4	23%
dropbox.com	244.0	2%
plus.google.com	231.2	2%
googleusercontent.com	163.1	1%
akamaiedge.net	141.7	1%
amazonaws.com	99.0	1%
wunderground.com	69.8	1%
live.com	39.2	0%
akadns.net	35.2	0%

Simple Network Management Protocol (SNMP)

- Byte and packet counters on individual links in the network topology
- Not information about individual flows
- Useful for monitoring utilization of links
- Little (no?) information about users



How This Data is Used By Research and Development

- NetFlow: developing intrusion detection, traffic engineering algorithms, studying user behavior
- DNS Traffic: Malware detection, performance optimization algorithms
- SNMP: (similar to NetFlow)
- DPI: Playback for testing of new forwarding algorithms, demand characterization, spam filtering, malware analysis, …

Summary

- ISPs, developers, and researchers use network data for research, development, security.
- Each type of data poses certain privacy risks.
- We need to understand how each type of data is used, and what the risks are, to make appropriate recommendations.